NASA TECHNICAL TRANSLATION NASA TT F-13,874

SEQUENTIAL CODING OF INFORMATION PRODUCED BY DISCRETE SOURCES WITH A FINITE MEMORY

V. N. Koshelev

N72-14217 (NASA-TT-F-13874) SEQUENTIAL CODING OF INFORMATION PRODUCED BY DISCRETE SOURCES WITH A FINITE MEMORY V.N. Koshelev (Techtran Corp.) Sep. 1971 10 p CSCL 09D

Unclas
12020
G3/08

Translation of: "Posledovatel'noye Kodirovaniye Informatsii, Proizvodimoy Diskretsmi Istochnikami s Konechnoy Pamyat'yu," IN: Chetverty Simpozium po Probleme Izvytochnosti v Informatsionnykh Sistemakh, (Fourth Symposium on the Problem of Redundancy in Information Systems), Leningrad, USSR Academy of Sciences, 1970, pp 378-387

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D. C. 20546 SEPTEMBER 1971

SEQUENTIAL CODING OF INFORMATION PRODUCED BY DISCRETE SOURCES WITH A FINITE MEMORY

V. N. Koshelev

ABSTRACT: A scheme of sequential coding of information generated by a discrete source of the Markov type is investigated. The coding can be carried out with an arbitrary transmission coefficient S, where S is greater than unity or less than unity, depending on whether or not it is necessary to "compress" the information or introduce additional redundancy to increase the reliability of transmission through the channel. The upper limits for the probability of error and the average number of operations in the case of sequential decoding are derived. An expression is found for the "computational entropy" of the source H_{comp}; it is shown that the average number of operations on a symbol is restricted to a constant which for $SH_{comp} < C_{comp}$, where C_{comp} is the "computational throughput" of the communication channel, does not depend on the value of the coding delay.

INTRODUCTION

/379*

This paper represents a generalization of well-known methods of sequential coding and decoding to the case where the source of information transmitted along the channel is described by a uniform Markov chain of order m. As the results show, sequential methods are suitable not only for coding at the channel's input and decoding at its output, but also for direct coding of the source's output into input signals of the channel, whereby decoding of the channel's output permits recovering the information directly in the same form in which it was generated by the source at the input.

In particular, if the channel is noiseless, then the proposed scheme can be used for statistical coding and decoding of sources which possess redundancy. In the general case the output generated by the source can be "compressed" in--sofar as the statistical properties of the source and the channel permit this,

---

*Numbers in the margin indicate pagination in the foreign text.

or, the output can be "spread out" as much as is necessary to increase the transmission's reliability.

Descriptions are given below of the coding and decoding algorithms, and the methods of obtaining upper limits to the average number of operations and the probable error produced upon reconstruction of the source's symbols at the channel's output are also discussed.

2. Fundamental Derivations /380

A *source of messages* generates a sequence of informational symbols $\{i_n\}$, $n = 0, \pm 1, \pm 2 \ldots$, each of which takes on values from some finite set $\{i\} = T$. We associate $|T|^m \times |T|^m$ with this source as the matrix of transition probabilities

$$P = \left[\mathcal{P}\{\alpha_1 \ldots \alpha_m | \beta_1 \ldots \beta_m\}\right], \quad \alpha, \beta \in \mathcal{T}; \qquad (4)$$

such that for $m \geqslant 2$

$$\mathcal{P}\{\alpha_1 \ldots \alpha_m | \beta_1 \ldots \beta_m\} = 0 \text{ for } (\alpha_1 \ldots \alpha_{m-1}) \neq (\beta_2 \ldots \beta_m)$$

and for $(\alpha_1 \ldots \alpha_{m-1}) = (\beta_2 \ldots \beta_m)$.

$$\mathcal{P}\{\alpha_1 \ldots \alpha_m | \beta_1 \ldots \beta_m\} = \mathcal{P}\{\alpha_m | \alpha_{m-1} \ldots \alpha_1 \beta_1\}$$

For fixed $1_{-m+1} \ldots L_0 I_1 \ldots 1_n$ we will use the concise notation

$$p(i_1 \ldots i_n) = \mathcal{P}\{i_1 \ldots i_n | i_{-m+1} \ldots i_0\} = \prod_{s=1}^{n} \mathcal{P}\{i_s | i_{s-m} \ldots i_{s-1}\}$$

The whole number m is called the source's memory.

A channel without a memory is determined by finite input and output alphabets $A = \{a\}$ and $B = \{b\}$ and by the transition probability matrix

$$Q = \left[q(b|a)\right], \quad a \in A, \; b \in B$$

A random output signal $a = (a_1 \ldots a_n)$, $a \in A$ corresponds to the input signal $b = (b_1 \ldots b_n)$, $b \in B$, such that

$$\mathcal{P}\{\bar{b}|\bar{a}\} = q(\bar{b}|\bar{a}) = \prod_{l=1}^{n} q(b_l|a_l).$$

It is assumed that the source's rate of operation relative to the channel can be arbitrarily varied in such a way that during the time when the source is generating a sequence consisting of $u \geqslant 1$ symbols, the channel has time to transmit a sequence consisting of $V \geqslant 1$ symbols. /381

Direct sequential coding of a message $i_1 i_2 \ldots$ into the input signal $a_1 a_2 \ldots$ for fixed symbols $i_{-1} i_0$ is specified with the help of the function H of ku and arguments for which sections of the message $i_{-(k-1)u+1} \ldots i_u$, $i_{-(k-2)u+1} \ldots i_{2u}$ and so on are used. The function can take on arbitrary values from the A^V space. Thus for any integral $\tau \geqslant 1$ some input signal

$$(a_{1,i_1 \ldots i_{su}} \ldots a_{sv,i_1 \ldots i_{su}}) = \bar{a}_{i_1 \ldots i_{su}} \quad \text{where} \quad \begin{aligned} a_1 \ldots a_v &= \mathcal{H}(i_u \ldots i_1 i_0 \ldots i_{-(k-1)u+1}) \\ a_{v+1} \ldots a_{2v} &= \mathcal{H}(i_{2u} \ldots i_{u+1} i_u \ldots i_1 i_0 \ldots i_{-(k-2)u+1}) \end{aligned}$$

is put into correspondence with the message $i_1 \ldots i_{su}$, etc. The elements of the spaces T_V^V and A^V are called, respectively, the message and signal blocks. The set of signal sequences $\{\bar{a}_{i_1 \ldots i_{sn}}\}$ possesses a tree-like structure and is called a tree with a length S blocks; $|T|^u$, consisting of blocks with a length of V symbols, proceeds from each point of a branch of the tree, the apex.

There corresponds to each block in a tree of length k its own set of arguments of the coding function. The quantity k is called the coding delay. The ratio $S = u/v$ of the message block's length to the signal block's length will be called the transmission coefficient; if $S < 1$, then the coding introduces an additional redundancy into the signals being transmitted; if $S > 1$, then the /382 source's representation in the coding tree cannot be unique, and this case corresponds to compression of the source, when a decrease in the redundant information in the signal occurs in comparison with that which is contained in the original message.

3. Decoding

Recovery of the original message from the signal obtained at the channel's output is accomplished by the method of decoding according to the maximum of the probability function calculated according to the combined distribution defined by the product { source × channel }.

We will discuss the effects of the decoding algorithm associated with the recovery of the first u symbols (i.e., of the first block) of the source. For convenience in writing, we will denote the message executed by the source by $O_1\ O_2 \ldots$, in contrast to the arbitrary messages $i_1 i_2 \ldots$

The quantity

$$\alpha^{sv} q(\bar{\beta}^{sv} | a_{i_1 \ldots i_{su}}) \, p(i_1 \ldots i_{su}) \tag{1}$$

will serve as the *criterion* for verifying a segment of the message $i_1 \ldots i_{su}$.

All the effects associated with the measurement and storage of this quantity will be discussed as a single elementary decoding operation. We will call an apex open if the criterion's value in it has already been calculated but there is at least one among the $||T|^u$ apexes which follow directly after it in which the criterion's value has not been determined. The remaining apexes /383 naturally, are broken up into closed ones and indeterminate ones.

Algorithm. Prior to the start of decoding the zero-point apex (from which $|T|^u$ of the first blocks proceeds) is open, since the criterion is arbitrarily set equal to unity in it. At the first decoding step the criterion in one of the apexes of the tree's first level is calculated; at the second step the maximum among all the open apexes is selected and the criterion is calculated in one of the apexes which directly follows after it. This procedure is continued until one of the apexes of the tree's k-th level first fails to be the maximum apex at some step. In this case the solution that the first block of the transmitted signal is equal to the first block of the branch leading to this apex of the k-th level is adopted. Accordingly the first u symbols of the source are determined with this. The subsequent search is continued according to this same rule, but now only in that part of the tree which proceeds from the decoded apex.

The complexity of decoding a single block is measured in a given case by the number of apexes $N_1(0_1 \ldots 0_{kui}\ b^{-kv};\ K)$ which are kept in the incorrect part of the tree proceeding from the zero-point apex. One should add unity to this number, since it corresponds to a check of the correct branch's first block.

4. Estimate of the Average Number of Operations and the Probability of an Error

Upper limit of the number $N_1(\ldots)$. We will consider a k-level tree, i.e., a tree composed of initial segments of k branches with a length of k blocks. /384

We will take the correct branch of this tree and find the maximum value of the criterion at its apexes. We will discuss the set of all apexes on which the value of the criterion is less than this minimum. Evidently not a single one of these apexes can turn out to be a maximum, i.e., cannot be selected as the starting point for subsequent explorations. This follows from the fact that at each step of the decoding there is only one apex among the set of all open apexes which belongs to the correct branch (the zero-point apex can, in particular, turn out to be it). Since before the start of each step, the decoding device is shifted to an open apex with the maximum value of the criterion, then only those apexes can be checked in an incorrect part of the tree for which

$$\alpha^{sv} q(\bar{b}^{sv}|\bar{a}_{i_1 \dots i_{su}}) p(i_1 \dots i_{su}) \geq$$
$$\geq \min_{t=\overline{0,k}} \left\{ \alpha^{sv} q(\bar{b}^{sv}|\bar{a}_{0_1 \dots 0_{su}}) p(0_1 \dots 0_{su}) \right\}, \tag{2}$$

and also those which directly follow after the apexes would satisfy the condition in equation (2), and on which this condition first turns out to be violated. Here and everywhere below one should assume that $i_1 \dots i_u \neq 0_1 \dots 0_u$.

Let

$$\lambda_{i_1 \dots i_{su}}(0_1 \dots i_{su}; \bar{b}^{sv}; K) = \begin{cases} 1, & \text{if equation (2) is fulfilled} \\ 0, & \text{if (2) is not fulfilled.} \end{cases} \tag{3}$$

Then

$$N_1(\cdot) \leq |T|^u \sum_{s=1}^{k} \sum_{i_1 \dots i_{su}} \lambda_{i_1 \dots i_{su}}(\dots) \tag{4}$$

The coefficient $|T|^u$ appears to be the first which violates the condition /385
of equation (2) because of the estimate of the number of apexes. Furthermore, following the reference [1], we will estimate the function in equation (3) as the quadratic root of the ratio of the left to the right parts of the inequality (2), we are free of the function's minimum in the denominator of this ratio, summing all the k + 1 ratios over $t = \overline{0_1 k}$, and we average the right part of the inequality (4) over the three probability ensembles, assuming them to be independent: over the message ensemble, the ensemble of noise in the channel, and the ensemble of all tree-like codes, determining all the coding functions with coding delay k and transmission coefficient S = u/v; it is assumed that an equilibrium distribution corresponding to some distribution $\tau(\alpha)$ over the input alphabet A is assigned in the last ensemble. As is well-known, an important condition is fulfilled in the case of such a distribution over an ensemble of tree-like codes. The condition consists of the fact that the

correct branch of the tree $a_{0_i \ldots 0_{tu}}$ and the branches of the incorrect part of the tree $a_{i_1 \ldots i_{su}}$, s,t = 1,k are mutually independent.

As a result we obtain for the average value of the quantity $N_1(\ldots)$ the estimate

$$\mathcal{N}_1 \leqslant \frac{|\mathcal{Y}|^{u}\left(z^{max}/z_1^{min}\right)^2}{\left(1-2^{-\lambda v[C(1)-SH(1)]}\right)\left(1-2^{-(1-\lambda)v[C(1)-SH(1)]}\right)} \quad (5)$$

(see below for the notation).

By means of similar discussions we will arrive at an estimate of the probability of an error (i.e., an event consisting of the fact that one of the branches of the tree's incorrect part turned out to have all k of the first apexes checked):

$$P_{calc} \leqslant \min_{\rho=\overline{0,1}} \frac{\left(z_\rho^{max}/z_\rho^{min}\right)^{1+\rho}}{1-2^{-\lambda v\rho[C(\rho)-SH(\rho)]}} 2^{-kv(1-\lambda)\rho[C(\rho)-SH(\rho)]} \quad (6)$$ /386

The following notation is used in the estimates of equations (5) and (6);

$$C(\rho)=\max_{\{r(a)\}} \frac{-\log \sum_{b}\left[\sum_{a} r(a)\, q(b|a)^{1/1+\rho}\right]^{1+\rho}}{\rho},$$

$$H(\rho)=\frac{\log \xi_\rho^{1+\rho}}{\rho},$$

Where ξ_ρ[1] is the largest Eigenvalue of the matrix P_ρ, which is obtained if all the elements are raised to the 1/1+ρ-th power in the source's stochastic matrix P, and Z_ρ^{max} and Z_ρ^{min} are, respectively, the maximum and minimum components of the eigenvector of matrix P_ρ, corresponding to the maximum eigenvalue ξ_ρ. We will call $H(1) = M_{calc}$ and $C(1) = C_{calc}$, respectively, the source's calculated entropy and the channel's calculated throughput. In such a case we have proved

[1]Symbol indicates eigenvalue.

the following

Theorem. If $S = u/v < C_{calc}/M_{calc}$, then there exists a coding function H with a coding delay k and a transmission coefficient S, such that the average number of decoding operations necessary for u symbols of the source satisfy the inequality (5), and the probability of an error satisfies the inequality (6), in the case of the weakest condition $S < C(o)/H(o)$, where

$$(1) \quad C(o) = \lim_{\rho \to 0} C(\rho) = \bar{C} \quad \text{and} \quad H(o) = \lim_{\rho \to 0} H(\rho) = H,$$

respectively, are the channel's throughput and the source's entropy. /387

In conclusion we note that the coding function K can be selected to be linear in the case of the derived limitations for the alphabets A and B.

REFERENCES

1. Gallagher, R., "A Simple Derivation of the Coding Theorem and Some Applications," *IEEE Tr. on Information Theory*, Vol. 11, No. 1, pp. 3-18, 1965.

Translated for the National Aeronautics and Space Administration under contract No. NASw-2037 by Techtran Corporation, P. O. Box 729, Glen Burnie, Maryland 21061, Translator: Thomas E. Margrave, Jr., Ph. D.